Policy Analysis Exercise

To Graduate from USAFA:

Identifying Which Admissions Criteria Make the Best Predictors of Cadet Success

Prepared by:
Beacher R. Webb III

Prepared for:
Rolland Stoneman,
Associate Director of Admissions
Kathleen O'Donnell,
Chief, Research and Assessment Division

Faculty Advisor: Alberto Abadie

Seminar Leader: Guy Stuart

April 2006

(submitted in partial fulfillment of the requirements for the degree of Master in Public Policy)

JOHN F. KENNEDY SCHOOL OF GOVERNMENT
HARVARD UNIVERSITY

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
01 APR 2006	**N/A**	

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
To Graduate from USAFA Identifying Which Admissions Criteria Make the Best Predictors of Cadet Success	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Harvard University**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited.

13. SUPPLEMENTARY NOTES
, The original document contains color images.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**UU**	**68**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

This page intentionally left blank.

EXECUTIVE SUMMARY

The purpose of this study is to identify which admissions criteria make the best predictors of cadet success at the United States Air Force Academy. Using graduation as the measure of cadet success, the study's findings and resulting recommendations seek to increase the efficiency of selection decisions and hence the graduation rate.

The study's methodology uses ten years of existing cadet data to quantify, rank, and interpret the admissions criteria's relationships to graduation. The Research and Assessment Division of the Academy's Plans and Programs Directorate supplied the data from its databases while withholding any information that could lead to personal identification.

The results of this study yield three types of information. First, they identify which admissions criteria have significant relationships with graduation and which do not. Table 1 and Table 2 summarize these findings. Second, the results lead to an ordinal ranking of admissions criteria in Table 3 based on the strength of their relationship to graduation. Third, the results provide a formula, shown in Figure 3, for determining an individual's probability of graduating from the Academy when given entries for each of the admissions criteria.

In addition to comparing individual admissions criteria, the study compares admissions composites currently used in the selection process. After finding the composites to be excellent predictors of graduation for the most part, the study looks for ways to improve the composites. It begins by comparing the strength of the academic and extracurricular composites' relationships to graduation and their current weights in the weighted composite. The study then evaluates how well the selection composite takes into account the relationships between individual admissions criteria and graduation. As a result, the study is able to provide recommendations for changing the weights of existing admissions criteria in addition to recommending a probability model for use as a new admissions criterion.

This report recommends the following:

- **Reduce the weight of or eliminate the admissions criteria without significant relationships to graduation listed in Table 2, such as student body leadership positions and Civil Air Patrol, and those with negative relationships at the bottom of Table 3, such as Academic Bowl and the Superintendent's Nomination.**

- **Increase the weight of the admissions criteria listed in Table 4, such as Eagle Scouts, National Honor Society, prior academic record, and having a parent or sibling who graduated from the Academy.**

- **Use the model presented in Figure 3 to determine future applicants' probability of graduating from the Academy. This probability can serve as a new admissions criterion or as a tool for calculating the total number of expected graduates in a class.**

This page intentionally left blank.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

The views expressed in this article are those of the author and do not reflect the official policy or position of United States Air Force, Department of Defense, or the U.S. Government.

INTRODUCTION

This study sets out to answer the question, "What admissions criteria make the best predictors of cadet success at the United States Air Force Academy?" This question is especially interesting, and perhaps difficult, given the Academy's uniqueness from other institutions of higher learning. Far from being the typical academically oriented college or university with the typical recruits looking for a college degree, the Academy hosts a program with an array of challenges outside the classroom in a mentally, emotionally, and physically demanding environment, and its student body are active duty, uniform-wearing service members with service obligations after graduation. As such the usual admissions criteria may not apply. In seeking candidates fit for the demands of the Academy and military service afterward, the Admissions Directorate considers a broader array of qualifications than those that merely appraise an applicant's academic ability, but which admissions criteria work best?

To answer this question, the study uses graduation as the measure of cadet success with the overarching goal of an answer being to increase the efficiency of selection decisions and hence the graduation rate. An efficient graduation rate is especially important considering that all cadets attend the Academy on full scholarships valued at approximately $320,000.[1] Accordingly, even small gains in the graduation rate from the study's findings could result in considerable savings as fewer appointments become necessary to achieve the same number of graduates. An even greater benefit is the ability to admit applicants better qualified to succeed at the Academy and in military service beyond.

[1] *Air Force Admissions*, 2005, <http://www.airforceadmissions.com/academy/>.

This page intentionally left blank.

BACKGROUND

The United States Air Force Academy mission is to "educate, train, and inspire men and women to become officers of character motivated to lead the United States Air Force in service to our nation."[2] Despite its uniqueness from other colleges, the Academy is still a prestigious, four year academic institution, the graduates of which hold accredited Bachelor of Science degrees. However, the four pillar philosophy at the Academy seeks to place equal emphasis on academics, military training, athletics, and character-building in a program dedicated at its core to officer development. In its mission statement, the Admissions Directorate states its role in the Academy mission:

> The Air Force Academy seeks young men and women who have demonstrated high levels of academic achievement as well as having sought out challenges and excelled, and who bring a diversity of talents, skills, viewpoints, and experiences - young people who meet the needs of the Air Force to include being pilot-qualified. We seek excellence in academics, leadership, athletics, and character. Racial, cultural, and gender diversity are valued. The first criterion for admission is, of course, the potential for success in our challenging academic and military training environment. We look for much more than just academic success. We seek leaders and people with exceptional character - people who possess the qualities and motivation to excel in our four-year experience and ultimately to go on and serve as warriors for our Air Force and nation.[3]

The admissions process begins with filling out a Pre-Candidate Questionnaire typically at or before the start of an applicant's senior year of high school. Upon meeting certain baseline standards of eligibility, the applicant becomes a candidate and must then compete for a nomination, the most common sources of which are the applicant's congressman or senator or the Vice President of the United States. Only after obtaining a nomination can a candidate become a nominee and so be eligible for consideration for an offer of appointment by the Admissions Directorate. In addition to the conventional standardized tests, transcripts, writing

[2] *Academy Admissions*, 2006, United States Air Force Academy, <http://www.academyadmissions.com/>.
[3] *Our Mission*, 12 Jan 2006, Admissions Directorate, United States Air Force Academy, <https://admissions.usafa.af.mil/rr/>.

sample, letters of recommendation, and application, the admissions process also includes a fitness test, medical and dental evaluations, and an interview.

The Admissions Directorate constructs composite scores from individual admissions criteria for evaluating nominees, as shown in Figure 1. It combines an applicant's highest standardized test scores with her prior academic record, a standardized version of high school GPA and class rank, to form an academic composite score. Meanwhile, an extracurricular composite score takes into account all athletic and non-athletic extracurricular activities pursued by the applicant. The Admissions Directorate then constructs a weighted composite from the academic and extracurricular composites. A Selections Panel considers all the remaining admissions criteria in assigning ratings that, along with the weighted composite, constitute the final selection composite. The Admissions Directorate then uses the selection composite to rank applicants in its selection decisions.

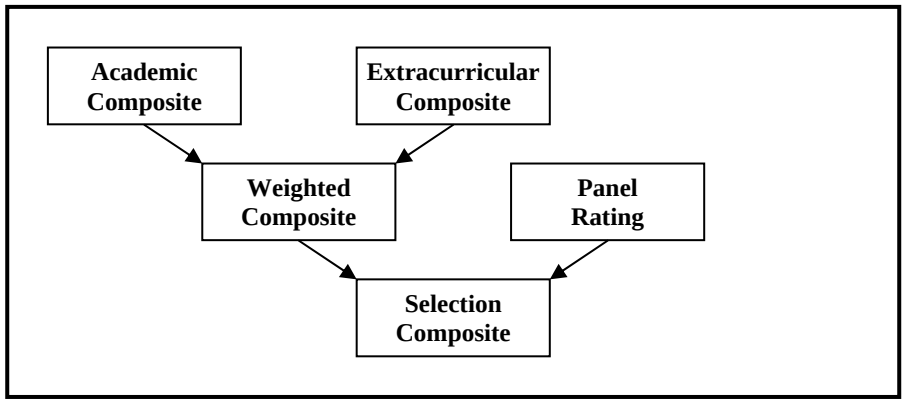

Figure 1. Construction of Admissions Composites

As indicated in the mission statement, the Admissions Directorate seeks to consider each admissions criterion and composite in accordance with its ability to indicate an applicant's "potential for success" at the Academy and in military service afterward.

METHODOLOGY

General: This study uses Probit regressions performed on ten years of cadet data to identify and quantify the relationships between admissions criteria and the likelihood of graduating from the Academy. The methodology consists of four phases: collecting the data, cleaning and formatting the data, performing and interpreting regressions, and reporting the results.

Data Collection: The Research and Assessment Division of the Academy's Plans and Programs Directorate supplied existing data from its databases on all cadets entering the Academy in the classes of 1996 through 2006. For each cadet, the data had entries for the admissions criteria considered in selection decisions, including the raw data on individual high school activities and test scores as well as the composite scores calculated from them with admissions formulas. For each cadet, entries also included cadet performance data, the available demographic information, and whether or not each individual graduated.

The data supplied by the Research and Assessment Division did not include names, social security numbers, or any other personal identifies. Because this study uses prior existing administrative data scrubbed of all personal identifiers, the Human Subjects Institutional Review Boards (IRB) at the Kennedy School of Government and USAFA have exempted it from human subjects review.

A critical part of this phase was working with the Admissions Directorate to pinpoint what the Academy is trying to maximize when it considers an applicant for admission. In order to answer the question, "What admissions criteria make the best predictors of cadet success at the USAF

Academy?" we had to first determine what measure of cadet success to use. While there are a plethora of cadet performance measures at the Academy, the Admissions Directorate identified a single outcome variable as its primary gauge of success, whether or not the cadet graduates. It should be noted that in its selection decisions the Admissions Directorate also seeks to maximize the performance of its selections beyond graduation in areas such as retention, advanced Professional Military Education selections, and promotion rates. As data becomes available on these topics, future studies could use the same methodology presented here to examine them.

Data Cleaning and Formatting: Data cleaning required screening the data for redundant entries and general mistakes, while data formatting involved converting the data into a format that could be recognized by the Stata statistical software used in this study. At times, this process required decisions to keep or delete certain data and how to best construct variables to represent certain admissions criteria. For a detailed account of the data cleaning and formatting process including the handling of redundant entries, the coding of binary variables, the selection of athletic variables, and the conversion of standardized test scores please see Appendix A.

Summary of the Data: Appendix B presents a complete list of all variable names and descriptions, and summary statistics for individual variables can be found in Appendix C. For a binary variable, the mean in Appendix C is the proportion of all cadets who fall in the category described by the variable. For example, the variable "graduated" has a mean of 0.736, indicating that 73.6% of the cadets entering these classes graduated.

The absence of dummy variables for the class of 1996, Caucasian as a race, and the state of Colorado as a place of birth is due to the fact that one category of each dummy variable set must be omitted so as not to over-constrain a regression. Each serves as a base case to which the other categories are compared. The class of 1996 was chosen as a base case, because it is the earliest year considered in the study. Caucasian was chosen as a base case, because most observations fall in that category. Colorado was chosen as a base case, because it is the home of the Air Force Academy.

In total, there are 12,406 observations for the cadets in the classes of 1996 through 2005, excluding international cadets. The database at the Academy is missing athletic and non-athletic activity data for the classes of 2000 and 2001, possibly due to the Academy's transition from its historic database to its current database. Therefore, the regressions requiring these variables used only the remaining eight years of data and their 10,078 observations. The purpose of using multiple years of data is only to control for time effects, which a regression should still be able to do satisfactorily with four years of continuous data on either side of the 2000-2001 gap.

There is also sibling data missing for the class of 2001. These data do exist in the Academy's historic database, but it is coded with a different set of Personal Identifier (PID) numbers than the remainder of the class of 2001 data taken from the current database. However, this missing data end up being irrelevant, because all the regressions in this study using sibling data also used athletic and non-athletic activity data and therefore already drop all data for the classes of 2000 and 2001 as discussed above.

Another set of variables missing data are the ratings from applicants' interviews with their liaison officers, known as LOCE interview ratings. The classes of 1999, 2002, 2003, and 2004 are missing these data entirely, and the remaining six classes are missing these data for up to a third of their members. As a result, this study is unable to analyze or control for the effects of LOCE interview ratings on graduation, a limitation that will be discussed more thoroughly later in this report.

The only other variables missing data are missing so few entries that it does not preclude their use in the study. The variable "par," a cadet's prior academic record based on high school GPA and class rank, is missing in 90 observations, less than one percent of the sample. Similarly, each of the admissions composites is missing four to eight entries, most of which overlap with those missing a prior academic record. As a result, the regressions using the 10,078 observations in the classes of 1996-1999 and 2002-2004 drop a total of only 89 observations missing data on one or more variables.

Regression Analysis: The simplest regressions use a linear model to estimate the expected value of a dependent variable given values for a set of explanatory variables. The coefficient of an explanatory variable in the model indicates how the value of the dependent variable is expected to change given a unit increase in the explanatory variable, while holding the other variables constant. In this way, the regression coefficients provide insight into the relationship between the explanatory variables and the dependent variable of interest.

The Probit regression model used in this study differs from the ordinary linear model in that it constrains predicted values of the dependent variable to fall between zero and one by using the cumulative distribution function of a standard normal distribution. Put simply, it treats the output of a linear model like a z-score. This constraint is especially useful for predicting the probability of binary dependent variables in that it accounts for the impossibility of probabilities greater than one or less than zero.

A second difference of the Probit regression model is that its explanatory variable coefficients are interpreted differently than those in a linear model. While a Probit coefficient shows the direction of a change in the dependent variable expected from a change in an explanatory variable, the magnitude of that change depends on the value of the explanatory variable and the values of all other variables. In order to compare explanatory variables based on the relative strength of their relationships with the dependent variable, this study reports marginal effects.

The dprobit feature of Stata performs the same Probit regression described above, but instead of displaying the resulting coefficients, it presents the marginal effect of increasing an explanatory variable from its mean when all other explanatory variables are held constant at their means. When the explanatory variable is a binary variable, the marginal effect given is for an increase from zero to one. As a result, the interpretation of marginal effects is similar to that of coefficients in an ordinary linear model regression. The dprobit feature still constructs all z-statistics, p-values, and confidence intervals from the Probit model coefficients and standard errors, so the interpretation of these statistics when determining statistical significance remains unchanged.

Seven regressions are performed in this study, and Appendix D summarizes their results. Regression 1 uses a Probit model and eight years of cadet data with graduation as the dependent variable and all admissions criteria and demographic controls as explanatory variables, excluding the admissions composites. The purpose of Regression 1 is to identify and quantify the relationships between the admissions criteria and a cadet's likelihood of graduating. This regression omits admissions composites, because they are constructed from the other criteria and, if included, would mask part or all of the other criteria's relationships with graduation.

In order to arrive at the final list of variables included in Regression 1, some method was necessary for deciding which variables to omit as clearly not statistically significant. While this report presents only seven regressions, much iteration preceded them to accomplish just this task. Initially, all variables indicated above were included in the regression. The first step was to cut 16 variables that were not statistically significant even at the 25% significance level. The regression was re-run and the process repeated to cut another eight variables at the 25% significance level. In the third round, the remaining eight variables not significant at the 10% significance level were eliminated. At this point, the regression was re-run adding each cut variable back into the regression one at a time to see if any had become significant at the 10% significance level as other variables were cut. None of the cut variables had become significant. The sets of dummy variables for class year, race, and place of birth were found to be jointly significant at the 10% significance level using the F-tests included in Appendix E, so they remained in the regression. They, and the other remaining significant variables, comprise the explanatory variables of Regression 1.

Regression 1 revealed that 18 admissions criteria, such as standardized test scores and high school varsity sports, along with four demographic controls have statistically significant relationships to graduation. These criteria are listed in Table 1 in the Results and Evaluation section of this report where they are discussed in great detail. The regression also revealed that 21 admissions criteria, such as student body leadership positions and Civil Air Patrol, along with one demographic control, age, do not have statistically significant relationships to graduation after controlling for those in Table 1.

Regressions 2-4 use a Probit model and all ten years of cadet data with graduation as the dependent variable. The only explanatory variables are the admissions selection composite in Regression 2, the admissions weighted composite in Regression 3, and the admissions academic and extracurricular composites in Regression 4. The purpose of these regressions is to identify and quantify the relationships between the admissions composites and graduation. These regressions revealed that the admissions composites have strong relationships to graduation.

Regression 5 is identical to Regression 1 except that it omits race and gender as explanatory variables. The coefficients from this regression resulted in a model for predicting the probability of graduation, which is presented in the Results and Evaluation section of this report.

Regression 6 uses an ordinary linear model and all ten years of cadet data with the admissions weighted composite as the dependent variable and the academic and extracurricular composites as the explanatory variables to provide a glance of the current weighting of the components of the weighted composite.

Regression 7 is identical to Regression 1 except that it adds the admissions selection composite as an explanatory variable. This regression seeks to identify how well the admissions selection composite takes into account all other predictors of graduation. If the current composition of the selection composite weights all other admissions criteria in a manner consistent with the strength of their relationship to graduation, then the inclusion of the selection composite should cause the coefficients of all other explanatory variables to become statistically insignificant. If another admissions criterion remains statistically significant, then its coefficient would suggest how the weighting of that criterion could be changed to strengthen the selection composite's relationship with graduation, and hence improve graduation rates. The findings from this regression are discussed at the end of the Results and Evaluation section.

All regressions in this study use robust standard errors as an added precaution against the possibility that another model might fit the data better than the Probit model.

Limitations: One limitation of this study is its inability to analyze or control for the relationship between LOCE interview ratings and the likelihood of graduation. If the study had included LOCE rating data, Stata would have dropped all observations missing these ratings, leaving a sample of only 5,242 observations. This outcome would not necessarily be problematic if the sample of cadets with LOCE ratings on file is very similar in composition to the sample of cadets missing the data. However, it is extremely unlikely that these two groups are consistently similar across all the variables considered. Indeed, without knowing why so many cadets *within* a class are missing the ratings, there is a possibility that missing ratings may be linked to one of the variables.

Taking the precaution of omitting the ratings entirely should have negligible effect on the regression results of all other variables, assuming there to be minimal omitted variable bias. In fact, it may be best to omit the ratings anyway if the interviewer bases his appraisals partially on other admissions criteria, because the ratings would then mask the relationships between these other criteria and graduation. For example, if a liason officer interviewer, being familiar with the interviewee's extracurricular activities, awarded the interviewee a higher leadership rating based on her participation in the Girls State program, then including the rating in the regression would then hide the importance of the Girls State program participation in predicting the likelihood of graduation.

A second limitation arises from the need to control for the effects of demographic variables such as race and gender, while also desiring to prevent discrimination in any proposed model based on these regressions. This topic is discussed in great detail in later sections of the report.

This page intentionally left blank.

RESULTS AND EVALUTION

General: The first five regressions of this study are designed to produce three types of information. First, the output identifies which admissions criteria have statistically significant relationships with the variable of interest, graduation, and which do not.

Second, the output leads to an ordinal ranking of admissions criteria based on the strength of their association with the variable of interest, graduation. In other words, regression output indicates which factors have the greatest effect on an individual's likelihood of graduation relative to other factors. Furthermore, for a specific individual, the output allows one to quantify each factor's impact on the individual's likelihood of graduation.

Third, regression output provides a formula that yields an individual's predicted probability of graduating from the Academy when given entries for each of the admissions criteria. This formula can be used to compare applicants based on their probability of graduating or to calculate the expected total number of graduates in a class.

The sixth and seventh regressions proceed to look for ways of improving the admissions composites' ability to predict graduation.

Identifying the Best Predictors of Graduation: Table 1 provides a list of the admissions criteria whose variable coefficients from Regressions 1-4 are found to be statistically significant. Effectively, Table 1 presents the admissions criteria found to have significant relationships with graduation. The asterisks indicate the standard of statistical significance met by each criterion.

A coefficient found significant at the 10% level means that if the regression was repeated many times with a different sample each time, the coefficient would appear different from zero due to random chance less than 10% of the time. Similar interpretations hold for the more conservative 5% and 1% significance levels. If follows that a coefficient significant at the 1% level is also significant at the 5% and 10% levels.

Table 1. Criteria With Statistically Significant Relationships to Graduation

Academic Criteria	
SAT or ACT score***	Prior academic record***
Extracurricular Criteria	
Number of high school varsity sports played**	Number of years spent playing high school varsity sports*
Being an Eagle Scout**	Participating in Academic Bowl*
Participating in Boys/Girls State***	Participating in a church group**
Participating in National Honor Society**	Participating in a student publication**
Other Criteria	
Candidate Fitness Assessment score***	Receiving a Superintendent's Nomination***
Having a parent who graduated from USAFA (Legacy)***	Having a parent who graduated from the U.S. Naval Academy*
Having a parent who graduated from the U.S. Military Academy*	Having a parent who graduated from the U.S. Coast Guard Academy*
Having a sibling who graduated from or is currently attending USAFA***	Having prior military service experience
Demographic Controls	
Gender*	Race*
USAFA class year at inprocessing***	Place of birth***

Notes: Prior service has statistical significance only at the 10.5% level.
* indicates statistical significance at the 10% level
** indicates statistical significance at the 5% level
*** indicates statistical significance at the 1% level

It should be noted that the coefficient on the variable "prior_service" is only statistically significant at the 10.5% level. The p-value for this coefficient wavered from just above to just below the 10% cutoff throughout the regression iterations, while none of the variable coefficients

cut from Regression 1 came close to 10% significance. As a result, "prior_service" was allowed to remain in the regression. The significance levels for race, class year, and place of birth are based on the F-tests of joint significance found in Appendix E.

Table 2, the counterpart to Table 1, lists those admissions criteria whose variable coefficients from Regression 1 are found not to be statistically significant at the 10% level. In other words, the admissions criteria in Table 2 are found not to have statistically significant relationships with graduation after controlling for the variables in Table 1.

Table 2. Criteria Without Statistically Significant Relationships to Graduation

Extracurricular Criteria	
Having a student body leadership position	Participating in Boy's/Girl's Nation
Participating in DECCA	Participating in Junior Achievement
Participating in 4H Club	Participating in a school club
Participating in Civil Air Patrol	Participating in a soaring program
Candidacy for a private pilot's license	Candidacy for a commercial pilot's license
Being an "All-State" athlete	Lettering in a high school sport
Being a team captain of a high school athletic team	Participating in community volunteer program
Participating in a non-athletic activity coded as "other"	Participating in a non-athletic activity coded as "unknown"
Other Criteria	
Attending the USAFA preparatory school	Receiving a Gold Award
Receiving a Falcon scholarship to another preperatory school	All nomination sources other than the Superintendent's Nomination
Having a parent who graduated from the U.S. Merchant Marine Academy	
Demographic Controls	
Age	

Some of the admissions criteria in Table 2 have statistically significant variable coefficients before controlling for the variables in Table 1, suggesting that the former are signals for the latter. For example, athletic criteria such as being an "All-State" athlete and lettering in a high school sport initially appear to have statistically significant relationships with graduation before controlling for the number of varsity high school sports played. This result suggests that being an "All-State" athlete and lettering in a high school sport are really just signals that the individual has played varsity sports, the criterion that actually has the relationship to graduation. Similarly, participating in a school club or having a student body leadership position are only significant before controlling for a student's prior academic record, and age is only significant before controlling for prior military service experience.

Taking a step further, Table 3 ranks the admissions criteria from Table 1 in order of their variable coefficients from Regressions 1-4, which is also an ordinal ranking by the strength of the variables' relationships to graduation. Class year, place of birth, and certain sports variables are not included in Table 3, but all are discussed later in this report. Each marginal effect reported comes from the dprobit regression results in Appendix D and is the marginal effect of increasing the corresponding explanatory variable from its mean when all other explanatory variables are held constant at their means. When the explanatory variable is a binary variable, the increase is from zero to one. With variables held constant at their mean, these marginal effects can be thought of as being those for an "average" cadet. Because of the Probit model used, the magnitude of the change in the dependent variable, "graduated," associated with a change in the listed explanatory variables depends on the level at which all other explanatory variables are held constant. Therefore, the marginal effects of each of the variables in Table 3

will vary from one individual to another. However, the ordinal ranking of the marginal effects will always remain the same, since the order is based on the relative magnitudes of regression coefficients, which do not change. For example, while the marginal effect of being an Eagle Scout will vary from one cadet to another, it will always be greater than the marginal effect of participating in Boys State.

Table 3. Ranking of Admissions Criteria by Strength of Relationship to Graduation

Variable Description	Variable Name	Marginal Effect (dF/dx)	Stan. Dev. of Variable	Impact on Probability of Graduating
Parent graduated from USCGA	parent_uscga	0.18887917		0.1889
Parent graduated from USNA	parent_usna	0.120379		0.1204
Sibling graduated from or is attending USAFA	sibling_usafa	0.117997		0.1180
Parent graduated from USAFA (Legacy)	parent_usafa	0.1087817		0.1088
Parent graduated from USMA	parent_usma	0.0888326		0.0888
Admissions Weighted Composite	**weighted_cmp**	**0.0015745**	**44.83791**	**0.0706**
Admissions Selection Composite	**sltd_cmp**	**0.0015142**	**46.51921**	**0.0704**
Admissions Academic Composite	**aca_cmp**	**0.0002373**	**288.366**	**0.0684**
Prior academic record	par	0.0006592	93.31395	0.0615
Race is African-American	black	0.0483123		0.0483
Eagle scout	eagle_scout	0.0400416		0.0400
Boys/Girls State participant	naa_bgstate	0.0354049		0.0354
National Honor Society member	naa_nhs	0.0270024		0.0270
Prior military service experience	prior_service	0.023275		0.0233
Admissions Extracurricular Composite	**leadership_cmp**	**0.0001004**	**183.7425**	**0.0184**
Candidate Fitness Assessment score	cfte_final_score	0.0001874	89.88607	0.0168
Composite of highest SAT or ACT scores	sat_act_conv	0.0001315	107.0976	0.0141
Most years spent playing a varisty sport	yrs_varsity	0.012402	0.6949036	0.0086
Church group participant	naa_church	-0.018869		-0.0189
Gender is female	female	-0.0224135		-0.0224
Student publication participant	naa_spub	-0.0261917		-0.0262
Academic Bowl participant	naa_acadbowl	-0.0317474		-0.0317
Superintendent's Nomination recipient	nom_sup	-0.1162281		-0.1162

Note: Marginal effects are taken from the dprobit coefficients of Regressions 1-4 (See Appendices C-G). The marginal effect given is the marginal effect of increasing an independent variable from its mean when all other variables are held constant at their means. When the independent variable is a binary variable, the increase is from zero to one.

In order to compare variables with very different scales, the marginal effect of each variable is multiplied by the variable's standard deviation to produce the impact in the far right column. For example, the marginal effect of "par" indicates a 0.000659 increase in the probability of graduating associated with a one point increase in a student's prior academic record, holding the other variables constant, while the marginal effect of "aca_cmp" indicates only a 0.000237

increase in the probability of graduating associated with an additional one point in the academic composite, holding the other variables constant. However, there is little use in comparing a unit increase in "par," which only varies from 400 to 809, to a unit increase in "aca_cmp," which ranges from 2406 to 4005. Instead, the values in the far right column allow one to compare the extrapolated impact of a one standard deviation increase in "par" to that of "aca_cmp," revealing that academic composite actually has a stronger relationship to graduation than does prior academic record. Standard deviations for binary variables are not included in Table 3, since their marginal effects and corresponding impacts are always interpreted as that associated with an increase in the variable from zero to one. Lastly, a horizontal bar in Table 3 separates those variables with positive relationships to graduation from those with negative relationships.

Interpreting the Results: This report now turns to discussing the results as they pertain to individual variables, starting with those at the top of Table 3 and working downward.

Parent graduation variables. The variables "parent_uscga," "parent_usna," parent_usafa," and "parent_usma" are binary variables indicating whether an individual had a parent graduate from one of the service academies. Traditionally, the admissions process awards some amount of "Legacy" points to candidates in such cases, and the results here seem to confirm the wisdom of the practice. Having a parent who graduated from the U.S. Coast Guard Academy, U.S. Naval Academy, U.S. Air Force Academy, or U.S. Military Academy is associated with 18.9, 12.0, 10.9, and 8.88 percentage point increases respectively in the probability of graduating from USAFA, holding the other variables constant at their mean. The larger figure for the USCGA could be prone to error, since there are only 18 cadets in the sample with a parent graduating

from there. Similarly, the fact that "parent_usmma" is found not to be statistically significant could be due to only 15 cadets in the sample having a parent who graduated from the U.S. Merchant Marine Academy. The number of cadets in the sample having a parent who graduated from the other academies is quite higher.

One possible explanation for these relationships is that children of parent graduates are better informed of the demands of life at an academy and in the military afterward and thus "know what they are getting into." Also, children of graduates may face greater pressure to follow in their parent's footsteps. Another possibility is that since all of these parents are college graduates, their children being more likely to graduate may reflect a more privileged upbringing or a heightened respect for education. Note that this study has no way of controlling for income.

Sibling graduation variable. The variable "sibling_usafa" indicates whether or not a cadet had a sibling who had already graduated from USAFA or was currently enrolled at the cadet's time of entry. Having such a sibling is associated with an 11.8 percentage point increase in the probability of graduating from USAFA, holding the other variables constant at their means. This phenomenon could again be a product of being better informed, or it could be the result of pressure not to fail where a sibling has succeeded. It could also be that siblings undertaking the rigors of the Academy together rely on each other for support, a factor amplified by the traditional, yet recently halted, practice of placing siblings in the same squadrons. If the latter is taken as the more accurate explanation, Training Group officials may want to revisit the policy of sibling placement. Whatever the case, "sibling_usafa" has one of the strongest relationships to graduation.

Admissions composite variables. The reported marginal effect of "sltd_cmp" shows that, on average, a one point increase in the selection composite is associated with a 0.151 percentage point increase in the probability of graduation, holding the other variables constant at their means. Extrapolating this result, an increase in the selection composite by one standard deviation is associated with a 7.04 percentage point increase in the probability of graduation. The weighted composite yields an almost identical result. These high figures should be extremely reassuring to the Admissions Directorate. As seen in Table 3, each of these composites is a better predictor of an individual graduating than nearly any other single admissions criterion.

The academic and extracurricular composites are expected to have somewhat lower individual impacts, since they jointly determine the weighted composite. However, the academic composite yields a result only slightly lower than that of the weighted composite, while the extracurricular composite's figure is much lower. An extrapolated one standard deviation increase in the extracurricular composite is associated with a 1.8 percentage point increase in the probability of graduation, holding the other variables constant at their means. The strength of the academic composite's relationship to graduation is more than three times that of the extracurricular composite, an interesting fact which this report returns to later.

Prior academic record. An extrapolated increase of one standard deviation in prior academic record is associated with a 6.15 percentage point increase in the probability of graduation, holding the other variables constant at their means. Comparing "par" to other variables in Table 3, this result suggests that prior academic record is a strong foundation of the academic composite for predicting graduation, which in turn is a strong foundation of the weighted

composite. Adding standardized test scores to the prior academic record to form the academic composite only improves its ability to predict graduation slightly. In the same way, adding the extracurricular composite to the academic composite to form the weighted composite only improves its ability to predict graduation slightly.

Race variables. While the race variables were jointly significant at the 10% level, only the variable "black" was significant independently with a p-value of 0.012. On average, African-American individuals have a 4.83 percentage point higher probability of graduation than Caucasian individuals, holding the other variables constant at their means. As in other cases, "holding the other variables constant" means comparing individuals who are identically qualified in all other respects. It is difficult to imagine an explanation. It could be that if discrimination is present at the Academy, African-Americans face it by hardening their resolve to graduate. On the other hand, if there is no discrimination at the Academy and substantial discrimination in the civilian labor market, then African-Americans may face strong economic incentives to graduate the Academy and enter the military rather than cope with reduced opportunities as a civilian.

Eagle Scout. Being an Eagle Scout is associated with a 4.00 percentage point increase in the probability of graduation, holding the other variables constant at their means. The variable "eagle_scout" has a stronger relationship with graduation than any other athletic or non-athletic activity. This result is likely due to the militaristic and well-rounded nature of the Boy Scouts of America program.

Boys/Girls State. Participation in Boys/Girls State is associated with a 3.54 percentage point increase in the probability of graduation, holding the other variables constant at their means. Interestingly, the variable coefficient for participation in Boys/Girls Nation is no longer statistically significant after controlling for participation in Boys/Girls State. This program likely signals propensity for public service and peer leadership.

National Honor Society. Participation in National Honor Society is associated with a 2.70 percentage point increase in the probability of graduation, holding the other variables constant at their means. Like Boys/Girls State, this program likely signals propensity for public service and peer leadership as well as academic competency because of its selection standards. While National Honor Society membership ranks behind Eagle Scouts and Boys/Girls State in its impact on the probability of graduation, it is a much more common program and may therefore prove a more useful tool for sorting applicants.

Prior Military Service. Having prior military service experience is associated with a 2.33 percentage point increase in the probability of graduation, holding the other variables constant at their means. Undoubtedly, having already completed basic training once successfully and spent some time in the military prepares an individual for the challenges faced at the Academy. In addition, these individuals may face substantial incentives to stay at the Academy and graduate, since many have military service obligations requiring them to return to enlisted status if they drop out.

Candidate Fitness Assessment Score. An extrapolated one standard deviation increase of 90 points in an applicant's Candidate Fitness Assessment score is associated with a 1.68 percentage point increase in the probability of graduation, holding the other variables constant at their means. It is likely that some of an individual's physical fitness is already captured by the athletic variables along with the teamwork and discipline associated with playing varsity sports, which may explain why this impact is relatively smaller than it might otherwise be. However, this variable remains statistically significant even well below the 1% significance level and has a meaningful, positive impact on the probability of graduation, which suggests that a high level of fitness prior to inprocessing may help one endure the physical strains of basic training and meet the fitness standards required of cadets.

Standardized Test Scores. An extrapolated one standard deviation increase of 107 points in an individual's highest SAT composite or converted ACT composite is associated with a 1.41 percentage point increase in the probability of graduation, holding the other variables constant at their means. There are a variety of explanations for why the impact of an individual's prior academic record is more than four times that of her standardized test scores. One is that a student's high school GPA and class rank may already reflect most of her academic ability or do so more accurately than the standardized tests. Another explanation is that the prior academic record captures a student's work ethic and study habits in addition to academic ability, virtues that the unique nature of the curriculum at the Academy may reward more than just academic ability alone.

Athletic variables. The athletic variable "num_varsity" is the number of high school varsity sports played. A binary variable simply indicating whether or not an individual participated in a varsity sport would not yield as much insight, since 97% of cadets in the sample played some varsity sport in high school. In fact, the average cadet played two or three. The regressions in this study use the quadratic variable "num_varsity2," which is just the square of "num_varsity," to account for the diminishing returns of additional sports played. For example, one would expect the impact on the probability of graduation to be greater for an increase in "num_varsity" from zero to one than for an increase from four to five, since the additional preparedness derived from playing one sport instead of none at all exceeds that from playing five sports instead of four. Because of the use of a quadratic variable, determining the impact on the probability of graduation requires calculating the probability of graduation for specific values of "num_varsity" using the Probit regression function from Regression 1. Figure 2 shows how the probability of graduation varies with the number of varsity sports played when all other variables are held constant at their means.

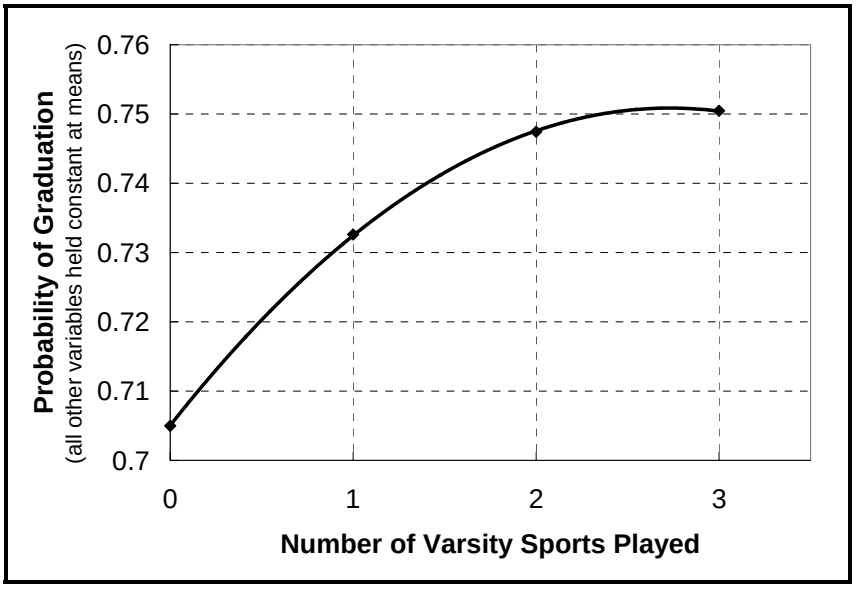

Figure 2. The Impact on the Probability of Graduation from Playing Sports

Playing one high school varsity sport instead of none increases a cadet's probability of graduation by 2.76 percentage points, holding the other variables constant at their means. The gain diminishes to 1.48 percentage points for a change from one sport to two and sinks to 0.30 percentage points when three sports are played instead of two.

One may desire to compare the strength of "num_varsity's" relationship to graduation with that of other variables. As mentioned previously, the ordinal ranking of variables in Table 3 by strength of relationship to graduation will not vary from one individual to another. This does not hold true for "num_varsity," so it is not included in Table 3. The quadratic term causes it to rank high among the admissions criteria in Table 3 for individuals who played zero or one sports but lower for those playing more sports.

The only other athletic variable to remain statistically significant after controlling for the number of varsity sports played was "yrs_varsity," which indicates the most years spent playing a single high school varsity sport. This variable only ranges from zero to three, because the Academy only records and considers sports played in the sophomore, junior, and senior years of high school. Together, the two variables "yrs_varsity" and "num_varsity" can account for many different combinations of athletic involvement, which is further facilitated by the fact that the Probit model interacts the two variables slightly. The results show that an extrapolated increase of one standard deviation in the most years spent playing a single varsity sport is associated with a 0.86 percentage point increase in the probability of graduation, holding the other variables constant at their means. Comparing the two athletic variables, playing another sport prepares an

applicant more than spending an extra year in a sport already played, suggesting that breadth of athletic experience may be more important than depth.

Church Group Participation. The discussion now turns to those variables found to have negative relationships with graduation. Participating in a church group prior to entering the Academy is associated with a 1.89 percentage point decrease in the probability of graduation, holding the other variables constant at their means. At first glance, this negative relationship seems strange, since one expects that as an additional venue for leadership and character building in the context of some moral framework, participation in a church body or youth group would improve an applicant's probability of graduation. However, cadets with religious backgrounds are much more prone to take advantage of the Academy's policy permitting a one or two year sabbatical for religious mission trips between the third and second class years. Following their mission trips, many cadets choose not to return to the Academy and incur no military service commitment, possibly explaining the lower graduation rate among individuals who participated in a church group. A second possibility is that cadets predisposed to a religious doctrine could be more likely to develop conscientious objections to military service and leave the Academy prior to incurring a service commitment.

Gender. On average, females have a 2.24 percentage point lower probability of graduating than males, holding the other variables constant at their means. As mentioned previously, "holding the other variables constant" means comparing individuals who are identically qualified in all other respects. It could be that gender discrimination, inter-gender hostility, or sexual assault hampers the graduation rate of females, though more benign explanations associated with gender

differences such as childbearing decisions and career choice preferences are certainly possible. It is important to note that if there is an atmosphere hostile to females to blame, then any effort by the Admissions Directorate to compensate by lowering admissions standards for females or using quotas may actually lead to a larger graduation gap due to low performance *and* exacerbate the stereotypes at the heart of the hostility, creating a cyclical negative effect.

Student Publication and Academic Bowl Participation. Participation in a student publication and participation in the Academic Bowl program are associated with 2.62 and 3.17 respective declines in the probability of graduation, holding the other variables constant at their means. Like participating in a church group, it is difficult to imagine how participating in these activities would make an individual less prepared for the Academy. However, it is possible that spending considerable time in these pursuits detracts from time spent studying or participating in a sport or other activity that would better prepare the individual for the Academy.

Nomination Source. As previously indicated, only the Superintendent's nomination was found to have a statistically significant relationship with graduation. On average, the probability of a Superintendent's nomination recipient graduating is 11.6 percentage points lower than recipients of other nominations, holding the other variables constant at their means. The Academy offers approximately 50 Superintendent nominations a year to well-qualified applicants who are unable to secure a nomination from another source. The result here suggests that these applicants' inability to obtain nominations from a primary source signals some ineptitude that hinders them from graduating.

Class Year Variables. The study includes class year variables to control for time effects on the graduation rate, such as changes to Academy policy and program difficulty over time. While the statistically significant coefficients on these variables and their marginal effects are of little use as admissions criteria, it is interesting to note how the probability of graduating varies from class to class even after controlling for admissions qualifications. For example, cadets in the class of 1997 and the class of 2005 had, on average, 5.79 and 3.76 percentage point lower probabilities of graduating respectively than did cadets in the class of 1996, holding the other variables constant at their means. In contrast, cadets in the class of 2002 and the class of 2003 had, on average, 13.8 and 11.3 percentage point higher probabilities of graduating respectively than did cadets in the class of 1996, holding the other variables constant at their means. Differences between the graduation probabilities in the other classes considered and that of the class of 1996 are not statistically significant.

Place of Birth Variables. While this set of demographic control variables is jointly significant, only the variable for data coded "unknown" and the variables for a birthplace in Florida, Montana, Tennessee, or Utah are statistically significant individually. Each of these five variables has a negative relationship with graduation, so that on average cadets born in the corresponding states have a lower probability of graduating than do cadets born in Colorado, the base case. The disparity is especially great for cadets with their place of birth coded "unknown." On average, the probability of such cadets graduating is 29.7 percentage points lower than that of their Colorado peers, holding the other variables constant at their means.

A Model for Predicting Probability: Up to this point, the discussion has focused on using the regression results to identify, quantify, and compare the relationships between admissions criteria and graduation. However, the regression results also provide a model for calculating an applicant's predicted probability of graduating from the Academy. The regression function itself serves as the formula that yields this probability. Such a formula can be used to compare applicants based on their probability of graduating or to calculate the expected total number of graduates in a class.

The formula presented in Figure 3 is the regression function for Regression 5, which is identical to Regression 1 except for the omission of race and gender. These two variables are left out in the interest of providing the Admissions Directorate with a model that does not discriminate based on race and gender. This decision is discussed in greater detail in the "Political Considerations" section of the report.

While the formula in Figure 3 appears intimidating, it is simple to program into a spreadsheet. The formula involves multiplying each variable included in Regression 5 by its coefficient and then finding the sum of the resulting products. There is an exception for class year variables. While these variables must be included in Regression 5 to control for time effects, the formula is intended for use with future applicants not belonging to the classes of 1996-2005. Therefore, each class year variable coefficient has already been multiplied by the variable mean and added to the regression constant. As a result, the regression constant changed slightly from -1.69 to -1.64.

P(Y=1 | sat_act_conv, par,…foreign_origin) = Φ(-1.63852 + 0.0003837sat_act_conv + 0.0020174par + 0.005928cfte_final_score + 0.136265eagle_scout + 0.4193941sibling_usafa + 0.3834645parent_usafa + 0.4381271parent_usna + 0.3024012parent_usma + 0.8462252parent_uscga – 0.3241165nom_sup + 0.1000076num_varsity – 0.0179985num_varsity2 + 0.0360439yrs_varsity – 0.0886263naa_acadbowl + 0.1172574naa_bgstate – 0.0597779naa_church + 0.0788085naa_nhs – 0.0852279naa_spub + 0.0873775prior_service – 0.803378unknown_origin – 0.2288518alabama + 0.0373825alaska – 0.0027327arizona – 0.1419045arkansas + 0.0251406california + 0.1413104connecticut + 0.1927075delaware + 0.013043dc – 0.190733florida – 0.0802206georgia – 0.1110898hawaii – 0.2516184idaho – 0.0404378illinois – 0.0665943indiana + 0.1328662iowa + 0.1195298kansas – 0.1858351kentucky + 0.0020205louisiana + 0.3844017maine + 0.1535083maryland + 0.1056386massachusetts – 0.034942michigan – 0.1173743minnesota – 0.1185046mississippi – 0.1484784missouri – 0.3563128montana + 0.2124126nebraska + 0.1039937nevada – 0.197433new_hampshire + 0.0052729new_jersey – 0.157743new_mexico + 0.0717923new_york – 0.1888925north_carolina – 0.1545421north_dakota + 0.1601866ohio – 0.0474061oklahoma – 0.1855306oregon + 0.0293215pennsylvania – 0.0981463puerto_rico – 0.2075892rhode_island – 0.0517633south_carolina + 0.28852south_dakota – 0.3305848tennessee – 0.0962686texas – 0.4554438utah + 0.2088133vermont + 0.0141811virginia + 0.1356272washington + 0.0765579west_virginia + 0.1138722wisconsin – 0.1660662wyoming + 0.0658606foreign_origin)

Figure 3. Model for Predicting Probability of Graduation

After plugging in the applicant's relevant value for each variable, the sum of all the terms included in the parentheses is a z-score for the applicant. The last step is to use this z-score in the standard normal cumulative distribution function, annotated Φ(z), to yield the predicted probability of graduation for the applicant. This function is given by:[4]

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} \exp\left(-\frac{z^2}{2}\right) dz$$

Many spreadsheets offer a feature pre-programmed with this function such as NORMSDIST( ) in Microsoft Excel.

[4] Alberto Abadie, "Binary Dependent Variables," Lecture Handout, John F. Kennedy School of Government, Cambridge, 22 Mar 2005.

With a probability calculated for each applicant, the Admissions Directorate could compare applicants using their probability of graduating as an additional admissions criterion. In addition, Admissions could find the total number of cadets expected to graduate from an entering class simply by summing the individual probabilities for all the members of the class. The Admissions Directorate may prefer to know the expected number of graduates in a new class during the selection process prior to offering appointments. Because the data used to construct the model includes only those applicants who accepted appointments and entered the Academy, the probability generated is the probability of an applicant graduating assuming that she is admitted and accepts her appointment. Therefore, the Admissions Directorate would simply need to multiply the sum of the probabilities of all applicants to be offered appointments by a constant reflecting the historical rate of appointment acceptance.

For example, suppose the Admissions Directorate intends to offer 1500 appointments and desires to know how many graduates to expect from the new class. Suppose further that from historical figures the Admissions Directorate expects a random 20% of these appointments to be rejected, resulting in the desired entering class size of 1200. Using the model yields a probability for each of the 1500 applicants offered appointments. Without knowing which of the 1500 will reject their appointments, the Admissions Directorate can sum the 1500 predicted probabilities and multiply the result by 80% to obtain the expected number of graduates from the class. Using this information, Admissions can then adjust the number of appointments offered to alter the expected number of graduates to match their target. To go a step further, the Admissions Directorate could use data on all applicants and a methodology similar to that in this report to investigate what types of applicants reject the appointments.

Improving Admissions Composites: In this section of the report, we begin to look for ways of improving the current admissions composites. We start with a brief glance at the current weights of the academic and extracurricular composites in constructing the weighted composite. From Regression 6 in Figure 4, the weighted composite is given by:

$$\text{weighted\_cmp} = 0.14\text{aca\_cmp} + 0.10\text{leadership\_cmp}$$

Here "weighted_cmp" is the weighted composite, "aca_cmp" is the academic composite, "leadership_cmp" is the extracurricular composite also know as the leadership composite, and the regression constant is approximately zero.

Source	SS	df	MS	Number of obs =	12399
				F(2, 12396) =	.
Model	24925399.4	2	12462699.7	Prob > F =	0.0000
Residual	9.87058226	12396	.000796272	R-squared =	1.0000
				Adj R-squared =	1.0000
Total	24925409.3	12398	2010.43792	Root MSE =	.02822

weighted_cmp	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
aca_cmp	.1399994	8.79e-07	.	0.000	.1399977	.1400011
leadership~p	.0999993	1.38e-06	.	0.000	.0999966	.1000021
_cons	.0031756	.0036677	0.87	0.387	-.0040136	.0103648

Figure 4. A Glance at the Components of the Weighted Composite in Regression 6

Based on the 14 to 10 ratio of the coefficients, the academic and extracurricular composites are weighted 58.3% and 41.7% respectively in the construction of the weighted composite. Recall, however, that the strength of the academic composite's relationship to graduation is more than three times that of the extracurricular composite in the case where all other variables are held constant at their means.

Therefore, if the Admissions Directorate desires to improve the weighted composite's efficiency as a predictor of graduation, and hence that of the selection composite as well, then there are two

general strategies. One is to weight the academic composite more heavily. The other is to improve the extracurricular composite.

Throughout the remainder of the report, this study generally favors the second strategy for three reasons. First, the fact that the ranking of the extracurricular composite's relationship with graduation in Table 3 is so far below that of the other composites and even certain admissions criteria used in building the extracurricular composite, suggests that there is significant room for improvement in the formula used to construct the extracurricular composite. Second, there is a limit to how well the academic composite can predict graduation by itself if the weight given to the extracurricular composite is allowed to become less and less significant. Third, concerns about maintaining the importance of all four pillars at the Academy and the well-roundedness they create in its officers may oppose weighting the academic composite more heavily, regardless of the potential efficiency gains in the graduation rate. Such concerns are addressed further in the "Political Considerations" section of this report. It should be noted that after improving the extracurricular composite, the relationships of each composite with graduation can always be revisited to see if the weights are still in need of changing.

The next step is to determine how the extracurricular composite ought to change, if the Admissions Directorate desires to improve its ability to predict graduation. As a starting point, those variables without statistically significant relationships to graduation in Table 2 and those with negative relationships in Table 3 make excellent candidates for a reduced weighting in the extracurricular composite or elimination altogether. Indeed, part of the reason the academic composite is such a strong predictor of graduation is that it is only composed of admissions

criteria with statistically significant positive relationships to graduation, prior academic record and standardized test scores.

Insights regarding the weighting of the remaining variables in Table 3 can be drawn from Regression 7. This regression is identical to Regression 1 except for the addition of the selection composite as an explanatory variable. If the current weighting of the selection composite adequately takes into account the strength of an admissions criterion's relationship to graduation, then the criterion's variable coefficient should no longer be statistically significant, and its marginal effect should shrink toward zero. If a variable coefficient remains positive and statistically significant but with a smaller marginal effect, then the weighting of the selection composite is only partially accounting for the admissions criterion's relationship to graduation. In this case, the selection composite would become a better predictor of graduation by weighting the criterion more heavily. If a variable is statistically significant with a negative coefficient and marginal effect, then the selection composite would become a better predictor of graduation by reducing that criterion's weighting.

Consider first the results from Regression 7 in Appendix D for admissions criteria found in the extracurricular composite. The coefficient of "yrs_varsity" is not statistically significant, indicating that its weighting in the extracurricular composite is adequate. In contrast, the marginal effect associated with the number of varsity sports played only fell from 0.032 in Regression 1 to 0.027 in Regression 7, indicating that this criterion is present in the final selection composite but not weighted heavily enough. The results yield similar conclusions for the Eagle Scout, Boys/Girls State, and National Honor Society variables, the marginal effects of

which only fell from 0.040 to 0.036, from 0.035 to 0.029, and from 0.027 to 0.021 respectively. Meanwhile, the negative coefficients on the Academic Bowl, church group, and student publication variables are consistent with the results in Table 3 in that they suggest the weights of these criteria should be reduced or eliminated altogether.

Looking next at those variables included in the academic composite, the variable coefficient for standardized test scores is no longer statistically significant in Regression 7, indicating an adequate weighting of the criterion in the selection composite. The marginal effect of prior academic record in Regression 7 is nearly half of that in Regression 1, showing that this criterion has substantial representation in the selection composite but could still benefit from a greater weight. To give prior academic record greater weight in the selection composite without affecting the weight of standardized test scores would require increasing the weight of prior academic record in the academic composite at the expense of standardized test scores and then increasing the weight of the academic composite in the weighted composite.

Finally, Regression 7 offers insight on other variables that pertain only to the selection composite. The marginal effects of the CFA score, prior service, and sibling and parent graduate variables all fall very slightly from Regression 1 and remain positive with statistically significant coefficients in Regression 7. The results imply that while a few points may be granted for these criteria, the selection composite would benefit from weighting them more heavily, especially the sibling and parent graduate criteria. Finally, the large negative marginal effect of the variable "nom_sup" supports penalizing the selection composite scores of recipients of the Superintendent's nomination.

This page intentionally left blank.

RECOMMENDATIONS

Based on the results presented here, this report recommends the following.

Reduce the weight of or eliminate the admissions criteria without statistically significant relationships to graduation listed in Table 2, such as student body leadership positions and Civil Air Patrol, and those with negative relationships at the bottom of Table 3, such as Academic Bowl and the Superintendent's Nomination. Reducing the weight given to any one or more of these variables in the construction of admissions composites should improve the composites' efficiency as predictors of graduation.

Increase the weight of the admissions criteria listed in Table 4 on the following page. Table 4 summarizes the admissions criteria that would benefit from heavier weighting based on the results of Regression 1, Regression 7, and the preceding discussion in the "Improving Admissions Composites" section of this report. Increasing the weight given to any one or more of these variables in the construction of admissions composites should improve the composites' efficiency as predictors of graduation.

Use the model presented in Figure 3 to calculate future applicants' predicted probability of graduating from the Academy. Once calculated, the Admissions Directorate can use these probabilities as an additional admissions criterion for comparing applicants or to determine the expected number of graduates from a class during the selection process as described in the "A Model for Predicting Probability" section of this report.

Table 4. Admissions Criteria Recommended for Additional Weighting

Criteria in the Extracirricular Composite	
Being an Eagle Scout	Participating in Boys/Girls State
Number of high school varsity sports played	Participating in National Honor Society
Criteria in the Academic Composite	
Prior academic record	
Criteria in the Selection Composite Only	
Candidate Fitness Assessment score	Having prior military service experience
Having a parent who graduated from USAFA	Having a parent who graduated from the U.S. Naval Academy
Having a parent who graduated from the U.S. Military Academy	Having a parent who graduated from the U.S. Coast Guard Academy
Having a sibling who graduated from or is currently attending USAFA	

POLITICAL CONSIDERATIONS

Each change to admissions criteria must be considered in light of the potential tradeoff between efficiency and the ideal qualifications of an officer. The Admissions Directorate's mission statement includes both seeking applicants with "potential for success" at the Academy and seeking "young people who meet the needs of the Air Force," though the two groups are not necessarily one and the same.[5] Each of the recommendations made in this report favor efficiency, seeking to change policy to admit the applicants most likely to graduate, but they do not account for what kinds of officers these applicants will become after graduation. The Admissions Directorate must weigh the benefits of an improved graduation rate against the potential costs of admitting "the wrong kind of officer candidate." For example, if SAT scores had been found to predict graduation perfectly, an efficiency argument would advocate offering appointments only to the applicants with the highest SAT scores, ignoring the fact that well-rounded applicants with leadership experience might make better officers.

Internal politics at the Academy and the four pillars philosophy typically keep this tradeoff in check. The Commandant of Cadets, Brig Gen Desjardins, would quickly object to the policy in the example above given her responsibility for the military and leadership development of the Cadet Wing, as would the Director of Athletics, Dr. Mueh, whose interest is in maintaining the physical fitness and leadership capacity of the Academy's graduates. Likewise, the Dean of the Faculty, Brig Gen Born, would resist a large-scale departure from academic criteria. The criteria recommended in this report for heavier weighting are well spread across the four pillars, as are those recommended for reduced weighting. The study also avoids changes to the weights of the academic and extracurricular composites in the weighted composite. As a result, the

[5] *Our Mission*.

recommendations presented here do not favor one pillar at the expense of another, which should minimize the internal resistance to a policy change should the Admissions Directorate decide on one.

Another relevant political topic is the necessity of avoiding discrimination on the basis of race and gender in selection decisions. This consideration pertains to the third recommendation proposing the use of an applicant's probability of graduation as an additional admissions criterion. The results of Regression 1 find that females are less likely to graduate than males and African-Americans are more likely to graduate than Caucasians after controlling for the effects of the other variables. Knowing that these relationships exist, if the model advocated in the third recommendation had included race and gender, then by using it Admissions would effectively take points away from female applicants and give points to African-American applicants. On the other hand, trying to be color-blind and gender-blind by coding all applicants the same race and gender would result in adding unearned points to female applicants and withholding earned points from African-American applicants. Each scenario is a different flavor of discrimination. In an effort to avoid either form of discrimination, the recommended model is based on Regression 5, which omits race and gender altogether. However, this does not solve the problem completely. Without controlling for these variables in the model, the effects of race and gender on graduation present themselves in other variables. For example, omitting gender results in the marginal effect of being an Eagle Scout, an all-male group, to increase from 0.040 to 0.042 so that extra points are still awarded to males indirectly. The only way to prevent discrimination completely is to not consider any admissions criterion that has performance or composition differences in race or gender, an impractical option for any admissions department.

CONCLUSION

Using a Probit regression model and ten years of existing cadet data, this study was able to identify the best predictors of cadet success at the United States Air Force Academy. After defining cadet success as graduation, the study quantified, ranked, and interpreted the admissions criteria's relationships to graduation. Furthermore, the study was able to provide a model for predicting an applicant's probability of graduating from the Academy. In addition to comparing individual admissions criteria, the study compared admissions composites currently used in the selection process. After finding the composites to be excellent predictors of graduation for the most part, the study proceeded to look for ways of improving the composites. As a result, the study was able to provide recommendations for changing the weights of existing admissions criteria in addition to recommending the probability model for use as a new admissions criterion.

This page intentionally left blank.

BIBLIOGRAPHY

"2000 SAT I-ACT Score Comparisons," *CollegeBoard.com*, 2006, The College Board, <http://www.collegeboard.com/sat/cbsenior/html/stat00f.html>.

Alberto Abadie, "Binary Dependent Variables," Lecture Handout, John F. Kennedy School of Government, Cambridge, 22 Mar 2005.

Academy Admissions, 2006, United States Air Force Academy, <http://www.academyadmissions.com/>.

Air Force Admissions, 2005, <http://www.airforceadmissions.com/academy/>.

Our Mission, 12 Jan 2006, Admissions Directorate, United States Air Force Academy, <https://admissions.usafa.af.mil/rr/>.

This page intentionally left blank.

APPENDIX A: Data Cleaning and Formatting Details

Data cleaning primarily consisted of reviewing the data in Microsoft Excel and deleting redundant entries and general mistakes. Most of the 96 cases of redundant entries involved two identical entries, so it did not matter which was deleted. In the other cases, one entry was clearly more current than the other based on the amount of information missing in the presumably older yet otherwise identical entry. In such cases, the older entry was deleted. During cleaning, all data for the class of 2006 were deleted, since graduation data for the class of 2006 will not be available until later this year. Also, international cadet entries were deleted due to the concern that their being subject to different admissions standards and processes could introduce "white noise" in the results.

Data formatting involved converting the data into a format that could be recognized by Stata, the statistical software package used in this study. The first step was to consolidate data from 19 source files into a single master file using Personal Identifier (PID) numbers and the Excel VLOOKUP function to match data. The second step required replacing the zero in each cell missing data with a blank so that Stata would read each as a missing entry rather than just a very low value. Next, binary variables such as graduating or being an Eagle Scout were converted into ones and zeros, with a one indicating a positive or success according to the variable name. For example, the variable "female," taking on a value of one for females and zero for males, replaced the database variable "Gender," coded "M" or "F." Likewise, sets of dummy variables replaced the categorical variables for class year, race, and place of birth.

The data supplied by the Research and Assessment Division included whether or not and for how many years each cadet was an All-State athlete, lettered player, team captain, or varsity player in each high school sport he played. All-State athlete, lettered player, and team captain status were each converted into a binary variable to reflect whether a cadet achieved that status in any sport at any time in high school. The last category, varsity player status, was initially broken down into a variable for each of the 22 sports on record so that sports could be analyzed individually. After conducting preliminary regressions, it seemed illogical that admissions decisions should discriminate based on the type of varsity sport played. Therefore, these variables were replaced with one for the total number of varsity sports played in high school, "num_varsity," and another indicating the most years spent participating at the varsity level in any one sport, "yrs_varsity."

Another challenge presented itself in the area of standardized test scores. Most cadets have either an ACT score on file or on SAT score on file, but not both. This could result from applicants choosing to take only one of the two tests or from the Admissions Directorate accepting only the higher of the two test scores. If the regression analysis included both tests, Stata would have dropped all cadets without scores for both tests on file, resulting in a very small, and biased, sample. If the regression analysis used only one test, Stata would have dropped all cadets who took the other test, and half the data would be lost. To get around this dilemma, ACT composite scores were converted to SAT composite scores using a table of equivalent score comparisons between the two tests based on percentile ranks and published by the College Board.[6] For individuals with scores for both tests on file, the higher score was taken. The result was a single variable reflective of the standardized test scores for each cadet, "sat_act_conv."

[6] "2000 SAT I-ACT Score Comparisons," *CollegeBoard.com*, 2006, The College Board, <http://www.collegeboard.com/sat/cbsenior/html/stat00f.html>.

APPENDIX B: Variable Descriptions

Variable	Description
graduated	graduated from USAFA
sltd_cmp	selection composite score
weighted_cmp	weighted composite score
aca_cmp	academic composite score
leadership_cmp	extracurricular composite score
sat_comp	SAT composite score
act_comp	ACT composite score
par	prior academic record score
cfte_final_score	candidate fitness assessment score
civil_air_patrol	Civil Air Patrol participant
eagle_scout	eagle scout
gold_award	gold award recipient
sibling_usafa	sibling graduated from or is attending USAFA
parent_usafa	parent graduated from USAFA
parent_usna	parent graduated from USNA
parent_usma	parent graduated from USMA
parent_uscga	parent graduated from USCGA
parent_usmma	parent graduated from USMMA
nom_afrotc	AFROTC nomination
nom_hms	honor military school nomination
nom_pres	presidential nomination
nom_reg_amn	regular airman nomination
nom_res_amn	reserve airman nomination
nom_sen	senatorial nomination
nom_sup	superintendent nomination
nom_vet	veteran nomination
nom_vp	vice presidential nomination
nom_other	other nomination
all_state	"all-state" high school athlete
lettered	lettered in a high school sport
team_captain	team captain of a high school athletic team
num_varsity	number of varsity high school sports played
yrs_varsity	most years spent playing a high school varisty sport
naa_other	other non-athletic activity participant
naa_4h	4H program participant
naa_acadbowl	Academic Bowl program participant
naa_bgnation	Boys/Girls Nation program participant
naa_bgstate	Boys/Girls State program participant
naa_church	church group program participant
naa_cpilot	commercial pilot license candidate
naa_decca	DECCA program participant
naa_ja	Junior Achievement program participant
naa_nhs	National Honor Society program participant

naa_ppilot	private pilots license candidate
naa_sclub	school club participant
naa_spub	school publication program participant
naa_soar	soaring program participant
naa_sbody	student body representative
naa_unk	participant in an unknown non-athletic activity
naa_commserv	community volunteer
loce_rtg_final	liason officer interview rating final score
loce_rtg_rec	liason officer interview rating for "recommend"
loce_rtg_selfconfid	liason officer interview rating for "self-confidence"
loce_rtg_humanrelat	liason officer interview rating for "human relations"
loce_rtg_planorg	liason officer interview rating for "planning / organizing"
loce_rtg_commskills	liason officer interview rating for "communication skills"
loce_rtg_leadership	liason officer interview rating for "leadership"
loce_rtg_motivation	liason officer interview rating for "motivation"
loce_rtg_preparation	liason officer interview rating for "preparation"
class_2005	entered USAFA with the class of 2005
class_2004	entered USAFA with the class of 2004
class_2003	entered USAFA with the class of 2003
class_2002	entered USAFA with the class of 2002
class_2001	entered USAFA with the class of 2001
class_2000	entered USAFA with the class of 2000
class_1999	entered USAFA with the class of 1999
class_1998	entered USAFA with the class of 1998
class_1997	entered USAFA with the class of 1997
female	gender is female
asian	self-identified race is Asian-American
black	self-identified race is African-American
hispanic	self-identified race is Hispanic
native	self-identified race is Native American
other_race	self-identified race is "other"
age	age in years
prior_service	prior military service experience
prep_usafa	graduate of the USAFA Preparatory School
prep_other	Falcon Scholar graduate of another preparatory school
unknown_origin	place of birth unknown
alabama	place of birth
alaska	place of birth
arizona	place of birth
arkansas	place of birth
california	place of birth
colorado	place of birth
connecticut	place of birth
delaware	place of birth
dc	place of birth
florida	place of birth

georgia	place of birth
hawaii	place of birth
idaho	place of birth
illinois	place of birth
indiana	place of birth
iowa	place of birth
kansas	place of birth
kentucky	place of birth
louisiana	place of birth
maine	place of birth
maryland	place of birth
massachusetts	place of birth
michigan	place of birth
minnesota	place of birth
mississippi	place of birth
missouri	place of birth
montana	place of birth
nebraska	place of birth
nevada	place of birth
new_hampshire	place of birth
new_jersey	place of birth
new_mexico	place of birth
new_york	place of birth
north_carolina	place of birth
north_dakota	place of birth
ohio	place of birth
oklahoma	place of birth
oregon	place of birth
pennsylvania	place of birth
puerto_rico	place of birth
rhode_island	place of birth
south_carolina	place of birth
south_dakota	place of birth
tennessee	place of birth
texas	place of birth
utah	place of birth
vermont	place of birth
virginia	place of birth
washington	place of birth
west_virginia	place of birth
wisconsin	place of birth
wyoming	place of birth
foreign_origin	place of birth

This page intentionally left blank.

APPENDIX C: Variable Summary Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
graduated	12406	0.735612	0.4410	0	1
sltd_cmp	12398	799.036900	46.5192	651	950
weighted_cmp	12399	623.373200	44.8379	483	763
aca_cmp	12400	3198.796000	288.3660	2406	4005
leadership_comp	12402	1755.330000	183.7425	768	2400
sat_comp	7386	1261.718000	110.9824	540	1600
act_comp	5036	28.989080	2.3857	20	36
par	12316	653.867800	93.3140	400	809
cfte_final_score	12400	500.832400	89.8861	200	800
civil_air_patrol	12406	0.082380	0.2750	0	1
eagle_scout	12406	0.072626	0.2595	0	1
gold_award	12406	0.001773	0.0421	0	1
sibling_usafa	11298	0.065675	0.2477	0	1
parent_usafa	12406	0.033855	0.1809	0	1
parent_usna	12406	0.003869	0.0621	0	1
parent_usma	12406	0.007658	0.0872	0	1
parent_uscga	12406	0.001773	0.0421	0	1
parent_usmma	12406	0.001290	0.0359	0	1
nom_afrotc	12406	0.009109	0.0950	0	1
nom_hms	12406	0.005481	0.0738	0	1
nom_pres	12406	0.076657	0.2661	0	1
nom_reg_amn	12406	0.023053	0.1501	0	1
nom_res_amn	12406	0.039900	0.1957	0	1
nom_sen	12406	0.170966	0.3765	0	1
nom_sup	12406	0.017492	0.1311	0	1
nom_vet	12406	0.002418	0.0491	0	1
nom_vp	12406	0.004917	0.0700	0	1
nom_other	12406	0.013139	0.1139	0	1
all_state	10078	0.385295	0.4867	0	1
lettered	10078	0.861084	0.3459	0	1
team_captain	10078	0.586327	0.4925	0	1
num_varsity	10078	2.369915	1.1380	0	10
yrs_varsity	10078	2.665311	0.6949	0	3
naa_other	10078	0.252332	0.4344	0	1
naa_4h	10078	0.011213	0.1053	0	1
naa_acadbowl	10078	0.082258	0.2748	0	1
naa_bgnation	10078	0.007839	0.0882	0	1
naa_bgstate	10078	0.209367	0.4069	0	1
naa_church	10078	0.676821	0.4677	0	1
naa_cpilot	10078	0.000794	0.0282	0	1

naa_decca	10078	0.009526	0.0971	0	1
naa_ja	10078	0.031455	0.1746	0	1
naa_nhs	10078	0.700040	0.4583	0	1
naa_ppilot	10078	0.052689	0.2234	0	1
naa_sclub	10078	0.490574	0.4999	0	1
naa_spub	10078	0.215519	0.4112	0	1
naa_soar	10078	0.004366	0.0659	0	1
naa_sbody	10078	0.444731	0.4970	0	1
naa_unk	10078	0.004862	0.0696	0	1
naa_commserv	10078	0.260270	0.4388	0	1
loce_rtg_final	5242	506.428800	111.8149	100	600
loce_rtg_rec	5049	4.738958	0.4685	2	5
loce_rtg_selfconfid	5049	4.715389	0.5039	2	5
loce_rtg_humanrelat	5049	4.667657	0.5248	1	5
loce_rtg_planorg	5049	4.768073	0.4538	2	5
loce_rtg_commskills	5049	4.569222	0.5669	1	5
loce_rtg_leadership	5049	4.704496	0.5030	1	5
loce_rtg_motivation	5049	4.804912	0.4488	2	5
loce_rtg_preparation	5049	4.582095	0.5978	1	5
class_2005	12406	0.101725	0.3023	0	1
class_2004	12406	0.106884	0.3090	0	1
class_2003	12406	0.106320	0.3083	0	1
class_2002	12406	0.097453	0.2966	0	1
class_2001	12406	0.089312	0.2852	0	1
class_2000	12406	0.098340	0.2978	0	1
class_1999	12406	0.106239	0.3082	0	1
class_1998	12406	0.103579	0.3047	0	1
class_1997	12406	0.092778	0.2901	0	1
female	12406	0.158472	0.3652	0	1
asian	12406	0.041109	0.1986	0	1
black	12406	0.056344	0.2306	0	1
hispanic	12406	0.069805	0.2548	0	1
native	12406	0.012575	0.1114	0	1
other_race	12406	0.003869	0.0621	0	1
age	12406	18.617160	0.7831	17	23
prior_service	12406	0.138643	0.3456	0	1
prep_usafa	12406	0.130904	0.3373	0	1
prep_other	12406	0.069402	0.2541	0	1
unknown_origin	12406	0.088828	0.2845	0	1
alabama	12406	0.010076	0.0999	0	1
alaska	12406	0.004756	0.0688	0	1
arizona	12406	0.017330	0.1305	0	1
arkansas	12406	0.006932	0.0830	0	1
california	12406	0.092617	0.2899	0	1
colorado	12406	0.039255	0.1942	0	1

connecticut	12406	0.007013	0.0835	0	1
delaware	12406	0.002660	0.0515	0	1
dc	12406	0.005481	0.0738	0	1
florida	12406	0.033129	0.1790	0	1
georgia	12406	0.020716	0.1424	0	1
hawaii	12406	0.006932	0.0830	0	1
idaho	12406	0.005401	0.0733	0	1
illinois	12406	0.034661	0.1829	0	1
indiana	12406	0.014912	0.1212	0	1
iowa	12406	0.013542	0.1156	0	1
kansas	12406	0.013945	0.1173	0	1
kentucky	12406	0.009834	0.0987	0	1
louisiana	12406	0.011446	0.1064	0	1
maine	12406	0.003789	0.0614	0	1
maryland	12406	0.014751	0.1206	0	1
massachusetts	12406	0.015638	0.1241	0	1
michigan	12406	0.024343	0.1541	0	1
minnesota	12406	0.024101	0.1534	0	1
mississippi	12406	0.008061	0.0894	0	1
missouri	12406	0.016686	0.1281	0	1
montana	12406	0.007819	0.0881	0	1
nebraska	12406	0.010559	0.1022	0	1
nevada	12406	0.005159	0.0716	0	1
new_hampshire	12406	0.003386	0.0581	0	1
new_jersey	12406	0.017089	0.1296	0	1
new_mexico	12406	0.009995	0.0995	0	1
new_york	12406	0.038127	0.1915	0	1
north_carolina	12406	0.017330	0.1305	0	1
north_dakota	12406	0.006207	0.0785	0	1
ohio	12406	0.038207	0.1917	0	1
oklahoma	12406	0.016282	0.1266	0	1
oregon	12406	0.009512	0.0971	0	1
pennsylvania	12406	0.031114	0.1736	0	1
puerto_rico	12406	0.002741	0.0523	0	1
rhode_island	12406	0.003789	0.0614	0	1
south_carolina	12406	0.009834	0.0987	0	1
south_dakota	12406	0.003063	0.0553	0	1
tennessee	12406	0.010801	0.1034	0	1
texas	12406	0.080123	0.2715	0	1
utah	12406	0.010721	0.1030	0	1
vermont	12406	0.001290	0.0359	0	1
virginia	12406	0.020232	0.1408	0	1
washington	12406	0.022247	0.1475	0	1
west_virginia	12406	0.004756	0.0688	0	1
wisconsin	12406	0.019910	0.1397	0	1

wyoming	12406	0.005723	0.0754	0	1
foreign_origin	12406	0.047155	0.2120	0	1

APPENDIX D: Summary of Regression Results

Dependent Variable: Graduated from USAFA (or weighted composite in regression 6)

Regression	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Admissions Composites							
selection composite score		.00151 ***					.00088 ***
		(.00009)					(.00020)
weighted composite score			.00157 ***				
			(.00009)				
academic composite score				.00024 ***		.14000 ***	
				(.00001)		(<.00001)	
extracurricular composite score				.00010 ***		.10000 ***	
				(.00002)		(<.00001)	
Academic Criteria							
SAT composite (ACT converted)	.00013 ***				.00012 **		-.00001
	(.00005)				(.00005)		(.00006)
prior academic record	.00066 ***				.00065 ***		.00038 ***
	(.00006)				(.00006)		(.00008)
Extracurricular Criteria							
eagle scout	.04004 **				.04206 ***		.03644 **
	(.01532)				(.01510)		(.01546)
number of varsity sports	.03207 **				.03210 **		.02686 **
	(.01254)				(.01260)		(.01261)
(number of varsity sports)^2	-.00580 ***				-.00578 ***		-.00566 ***
	(.00205)				(.00207)		(.00206)
most years in a varsity sport	.01240 *				.01157		.00851
	(.00733)				(.00731)		(.00736)
Academic Bowl	-.03175 *				-.02912 *		-.03592 **
	(.01809)				(.01798)		(.01824)
Boys/Girls State	.03540 ***				.03677 ***		.02853 **
	(.01107)				(.01102)		(.01132)
church group	-.01887 **				-.01905 **		-.02189 **
	(.00957)				(.00956)		(.00957)
National Honor Society	.02700 **				.02555 **		.02085 *
	(.01088)				(.01083)		(.01089)
student publication	-.02619 **				-.02778 **		-.02940 ***
	(.01127)				(.01122)		(.01133)
Other Criteria							
candidate fitness test score	.00019 ***				.00019 ***		.00017 ***
	(.00005)				(.00005)		(.00005)
sibling graduated USAFA	.11800 ***				.11745 ***		.11451 ***
	(.01447)				(.01444)		(.01465)
parent graduated USAFA	.10878 ***				.10747 ***		.10405 ***
	(.02057)				(.02060)		(.02100)
parent graduated USNA	.12038 *				.11876 *		.11726 *
	(.05535)				(.05603)		(.05587)
parent graduated USMA	.08883 *				.08689 *		.08592 *
	(.04395)				(.04409)		(.04445)
parent graduated USCGA	.18879 *				.18898 *		.18707 *
	(.06250)				(.06237)		(.06327)
superintendent nomination	-.11623 ***				-.11354 ***		-.10650 ***
	(.03693)				(.03675)		(.03676)
prior service experience	.02328				.02744 **		.02562 *
	(.01409)				(.01341)		(.01396)
Demographic Controls Used							
class year	yes ***	no	no	no	yes ***	no	yes ***
gender	yes *	no	no	no	no	no	yes **
race	yes *	no	no	no	no	no	yes *
place of birth	yes ***	no	no	no	yes ***	no	yes ***
R-squared	.068	.022	.022	.023	.067	1.000	.069
Number of observations	9989	12398	12399	12400	9989	12399	9988

Notes: All Probit regressions report marginal effects with robust standard errors in parentheses. Regression 6 reports OLS coefficients. Class year, race, and place of birth significance based on F-tests of joint significance.

* indicates statistical significance at the 10% level

** indicates statistical significance at the 5% level

*** indicates statistical significance at the 1% level

This page intentionally left blank.

APPENDIX E: Regression 1 F-Tests

```
( 1)  class_2005 = 0
( 2)  class_2004 = 0
( 3)  class_2003 = 0
( 4)  class_2002 = 0
( 5)  class_1999 = 0
( 6)  class_1998 = 0
( 7)  class_1997 = 0

          chi2(  7) =   172.28
        Prob > chi2 =     0.0000
```

Class Year F-Test

```
( 1)  asian = 0
( 2)  black = 0
( 3)  hispanic = 0
( 4)  native = 0
( 5)  other_race = 0

          chi2(  5) =    10.23
        Prob > chi2 =     0.0691
```

Race F-Test

```
( 1)  unknown origin = 0
( 2)  alabama = 0
( 3)  alaska = 0
( 4)  arizona = 0
( 5)  arkansas = 0
( 6)  california = 0
( 7)  connecticut = 0
( 8)  delaware = 0
( 9)  dc = 0
(10)  florida = 0
(11)  georgia = 0
(12)  hawaii = 0
(13)  idaho = 0
(14)  illinois = 0
(15)  indiana = 0
(16)  iowa = 0
(17)  kansas = 0
(18)  kentucky = 0
(19)  louisiana = 0
(20)  maine = 0
(21)  maryland = 0
(22)  massachusetts = 0
(23)  michigan = 0
(24)  minnesota = 0
(25)  mississippi = 0
(26)  missouri = 0
(27)  montana = 0
(28)  nebraska = 0
(29)  nevada = 0
(30)  new hampshire = 0
(31)  new jersey = 0
(32)  new mexico = 0
(33)  new_york = 0
(34)  north_carolina = 0
(35)  north dakota = 0
(36)  ohio = 0
(37)  oklahoma = 0
(38)  oregon = 0
(39)  pennsylvania = 0
(40)  puerto rico = 0
(41)  rhode island = 0
(42)  south carolina = 0
(43)  south_dakota = 0
(44)  tennessee = 0
(45)  texas = 0
(46)  utah = 0
(47)  vermont = 0
(48)  virginia = 0
(49)  washington = 0
(50)  west virginia = 0
(51)  wisconsin = 0
(52)  wyoming = 0
(53)  foreign_origin = 0

          chi2( 53) =    289.29
        Prob > chi2 =     0.0000
```

Place of Birth F-Test

This page intentionally left blank.

JOHN F. KENNEDY SCHOOL OF GOVERNMENT
HARVARD UNIVERSITY